

# Résultats des enquêtes sur la virescence florale du cotonnier effectuées en Haute-Volta de 1970 à 1978

R. DELATTRE (1) et A. JOLY (2)

## RÉSUMÉ

Des enquêtes régionales sur les champs de cotonniers atteints de virescence florale (VFC) ont été conduites pendant neuf années en Haute-Volta. Après examen critique des données recueillies, les fichiers ainsi constitués ont été exploités par deux programmes informatiques différents : régression multiple (R.M.) et analyse par segmentation (A.S.).

Les principaux facteurs analysés sont : date de semis ou de levée, dimension des champs, nombre de traitements, fumures, précédent cultural, nature des produits insecticides.

Les résultats sont énoncés par :

- des tableaux de fréquences de divers facteurs, simples ou croisés ;
- des formules de régression sur facteurs quantitatifs
- une étude des facteurs qualitatifs ;
- un arbre de classification obtenu par segmentation.

Seule la variable résultante des « taux de plants à l'hectare » est exploitable, celle du rendement n'est pas étudiable.

Des relations intéressantes s'établissent entre le taux de maladie et les trois premiers facteurs cités, en R.M. Celles mises en évidence par A.S. comportent des facteurs supplémentaires et apportent aussi des nuances complémentaires aux résultats élucidés en R.M. Le classement par ordre d'importance n'est pas le même dans l'une et l'autre analyses ; pour des années exceptionnelles, l'A.S. montre que l'action « dimension du champ » est modulée différemment quand la date des semis est tardive.

L'intérêt ou les inconvénients respectifs des deux approches statistiques sont évoqués ; on souligne l'éventuelle complémentarité des deux techniques d'études, expérimentation classique et collecte de données par enquêtes régionales ; ce dernier processus permet d'aborder certains problèmes (relation avec la dimension des champs, par exemple) qui ne peuvent être traités en expérimentation courante.

Quant aux facteurs étudiés en Haute-Volta concurremment par les deux techniques, il y a bonne concordance dans les conclusions, notamment à propos de l'incidence de la date de semis précoce sur le taux de la maladie. Néanmoins, la recommandation pratique d'effectuer des semis précoces est maintenue car elle est le fondement d'un bon potentiel de productivité tant que le taux de virescence n'est pas exceptionnellement élevé.

**Mots clés :** virescence florale, enquête épidémiologique, Haute-Volta, régression multiple, analyse par segmentation, date de semis et de levée, dimension des champs, applications insecticides, fumure minérale, précédent cultural, méthodes d'analyse.

## INTRODUCTION ET DONNÉES GÉNÉRALES

D'année en année, depuis neuf ans, des fiches d'enquête sur la virescence florale du cotonnier (VFC) ont été collectées en Haute-Volta, notamment dans les zones cotonnières les plus touchées par cette maladie.

Le but de cette opération était d'obtenir des informations d'ordre général sur la distribution de la maladie aux champs, les variations dans l'incidence régionale ou annuelle, les relations éventuelles avec la nature du sol ou divers facteurs culturels (semis, sarclage, fumure, précédent cultural, nombre de

traitements...). Nous ne pouvions nous lancer dans une enquête globale et détaillée aboutissant à l'évaluation de l'incidence économique à l'échelon national, qui eût exigé des moyens plus importants et des précautions statistiques beaucoup plus minutieuses ; nous n'avons recherché que la mise en évidence de relations entre facteurs d'ordre technique pouvant s'établir dans des champs contaminés à des degrés divers, pour les mettre en parallèle, si possible, avec les résultats expérimentaux obtenus à la ferme de Boni et dans notre réseau régional d'essais, relatés dans les rapports annuels de 1965 à 1978, en vue d'asseoir les conclusions sur des bases élargies.

Bon nombre des fiches renvoyées au centre de traitement informatique étant incomplètement rem-

(1) Division phytosanitaire, I.R.C.T., Paris.

(2) Service de biométrie et informatique, I.R.C.T. Montpellier.

plies, il a fallu d'abord les trier et constituer des fichiers d'effectifs variables selon le nombre de critères correctement indiqués.

Pour la revue la plus générale, on a retenu 823 fiches, qui se répartissent essentiellement entre les ORD\* de Bobo, Dédougou et Diébougou, la zone de Ouagadougou ne contribuant que pour 23 cas et seulement pour deux années. La répartition par année est variable, et le nombre de fiches collectées est faible en 1972 (37 fiches) et en 1978 (29 fiches) (tabl. 1).

Quant aux surfaces échantillonnées, elles représentent 1 227 hectares au total, s'étageant de 14 hectares en 1978 à 306 hectares en 1971.

Au total, le taux moyen de contamination est de 159 plants à l'hectare; les chiffres extrêmes, par année et région (Ouagadougou exclus), sont de 360 en 1974, à Dédougou, et de 71 en 1973, à Dédougou encore (tabl. 2).

L'analyse au niveau du secteur agricole (tabl. 3) est rendue très difficile par l'inégalité des effectifs des fiches dûment remplies. Ainsi, pour l'ORD de Diébougou, seule l'année 1976 peut fournir des indications très relativement valables sur l'intensité de la maladie: 206 plants à l'hectare à Dédougou centre, puis 167 à Dano, 105 à Dissin, 93 à Kampti, 44 à Nako et 20 à Gaoua (à Baïé on obtient 33 plants/ha, mais sur un seul échantillon).

De même, pour l'ORD de Ouagadougou, l'année 1977

est la seule où les divers secteurs ont fourni quelques indications, l'année 1973 étant l'autre contributrice pour le reste des informations par le secteur de Manga seulement. Les deux zones fondamentales pour la collecte de renseignements utilisés dans cette revue sont donc les ORD de Bobo et de Dédougou: encore y rencontre-t-on beaucoup de lacunes, comme le montre le tableau 3.

En moyenne, les secteurs de Boromo et de Houndé sont les plus touchés (310 et 233 plants atteints à l'hectare), tandis que Tougan et N'Dorola présentent des densités de 17 et 23 plants virescents à l'hectare. Les variations annuelles globales sont imputables essentiellement aux observations faites dans ces deux zones: les années 1973 et 1977 ont été peu marquées par la maladie (104 et 100 plants/ha), qui aurait prévalu notamment en 1972 (278 plants/ha) et en 1970 (203 plants/ha), puis en 1974 et 1975 (184 et 193 plants/ha). On notera surtout qu'il n'existe aucune tendance suivie quant à un accroissement de la sévérité de l'attaque d'année en année, mais seulement des oscillations irrégulières pendant la période considérée, l'amplitude relative allant du simple au triple environ (100 à 278 plants/ha); ceci est d'autant plus assuré que l'enquête a porté fondamentalement sur les champs ayant été repérés comme étant ou pouvant être les plus fortement infestés, plutôt que sur un échantillonnage complet des champs atteints ou non de la maladie.

## ÉTUDE DES VARIABLES ET DE LEUR DISTRIBUTION

Le programme d'histogramme appliqué à chacune des variables répartit les effectifs en 14 classes d'égale amplitude et en donne les fréquences relatives.

### Exploitation du fichier brut et complet (823 fiches)

Les résultats concernant la surface (en ares) des champs de cotonniers observés montrent que 19 % de ceux-ci ont une surface de moins de 3/4 d'hectare, 40 % sont d'une superficie comprise entre 0,75 et 1,4 ha, puis 25 % entre 1,4 et 2 ha, les surfaces cultivées de plus de 3,3 ha représentant moins de 5 % des observations.

Quant à la densité de la maladie, on montre que 76 % des champs ont entre 0 et 170 pieds atteints à l'hectare, puis 12 % entre 170 et 340 pieds, 5 % entre 340 et 510 pieds, quelques cultures présentant encore des valeurs supérieures à 1 000 pieds à l'hectare.

### Exploitation du fichier restreint (503 fiches)

a - Fréquence des surfaces des champs (en ares):

Taux en %	Surface (ares)	$\sqrt{\text{ares}} \times 100$
3,8	de 10 à 24	de 31 à 50
1,2	25 49	51 70
15,5	50 80	71 90
35,5	81 119	91 109
16,3	120 164	110 128
13,5	165 216	129 147
4,4	217 279	148 167
3,0	280 346	168 186
1,4	347 424	187 206
etc.		

Moyenne générale: 130,9 ares.

b - Les données sur la densité de l'attaque ont été transformées par « racine carrée », opération destinée à normaliser la distribution de cette variable issue de dénombrements et très dissymétrique. On obtient

\* ORD: Office régional de Développement.

alors un meilleur étalement sur les 14 classes que dans le cas de l'exploitation directe du fichier complet. Cette transformation sera adoptée dans l'étude en régression multiple qui sera faite ultérieurement.

c - D'autres histogrammes, fournis automatiquement pour les autres variables introduites en vue de l'analyse des corrélations, sont résumés en tableaux ci-après, qui donnent la répartition des fréquences relatives à des données calendaires simples.

#### Dates de première pluie :

1,7 %	avant le 70° j
3,8 %	entre le 71° et le 92° j
7,7 %	» le 93° et le 115° j
24,7 %	» le 116° et le 137° j
34,2 %	» le 138° et le 160° j
24,0 %	» le 161° et le 182° j
3,4 %	» le 183° et le 205° j, etc.

#### Dates de semis :

1,8 %	avant le 134° j
4,6 %	entre le 135° et le 144° j
8,7 %	» le 145° et le 150° j
16,9 %	» le 151° et le 156° j
24,7 %	» le 157° et le 163° j
8,9 %	» le 164° et le 175° j
7,7 %	» le 176° et le 182° j
4,0 %	» le 183° et le 188° j, etc.

#### Dates de levée :

2,7 %	avant le 148° j
6,0 %	entre le 149° et le 156° j
19,5 %	» le 157° et le 164° j
26,0 %	» le 165° et le 172° j
21,5 %	» le 173° et le 180° j
9,3 %	» le 181° et le 188° j
5,2 %	» le 189° et le 196° j, etc.

#### Dates de premier sarclage :

13,9 %	entre le 140° et le 157° j
41,5 %	» le 158° et le 175° j
30,2 %	» le 176° et le 192° j
11,1 %	» le 193° et le 209° j
2,2 %	» le 210° et le 227° j, etc.

d - Déduts de ces histogrammes pour « dates calendaires », les histogrammes relatifs à des « durées » entre premières pluies, semis, levée, premier sarclage, etc., peuvent être obtenus aisément ; comme ils sont d'une plus grande valeur indicative de la succession des travaux agricoles, soit en fonction des phénomènes météorologiques, soit par rapport à la séquence des interventions humaines, nous les relatons ci-après.

#### Intervalle entre première pluie et levée :

On met ici en évidence l'influence des facteurs météorologiques sur les processus physiologiques de germination, une fois assurée la mise en place des graines dans les poquets.

#### Fréquence (%)

34,2	moins de 12 j
15,9	entre 13 et 24 j
18,1	» 25 » 37 j
11,3	» 38 » 49 j
9,1	» 50 » 61 j
3,6	» 62 » 74 j
1,6	» 75 » 86 j
1,8	» 87 » 99 j
4,4	plus de 100 j

Il est probable que les chiffres des dernières catégories se rapportent à des intervalles entre le premier semis et une date de levée de semis ultérieurs, ou encore à des premières pluies très précoces, c'est-à-dire tombées bien avant l'époque consacrée normalement aux semis.

#### Intervalle entre semis et premier sarclage :

On notera que 75 % des champs sont sarclés moins de 35 jours après le semis, mais que le premier chiffre est douteux.

#### Fréquence (%)

7,1	moins de 11 j
29,2	entre 11 et 22 j
38,8	» 23 » 34 j
14,7	» 35 » 45 j
5,2	» 46 » 56 j
3,6	» 57 » 68 j
1,4	après 69 j

#### Intervalle entre première pluie et premier sarclage :

Il pourrait se déduire des données précédentes, mais son intérêt « hybride » est moins évident que celui de chacun de ces composants, car cet intervalle ne représente pas forcément la durée d'enherbement du champ, celui-ci étant une première fois nettoyé avant le semis.

#### Nombre d'applications insecticides :

On résume ainsi les données pour la période envisagée :

Fréquence (%)	Nombre d'applications
1	néant
3,6	1
17,7	2
41,6	3
30,2	4
5,4	5
0,4	6
0,2	7

Ce tableau ne tient pas compte de l'évolution de la technique des traitements au cours de la décennie ; on peut admettre qu'actuellement (1979-80), plus de la moitié des champs recevraient au moins 4 traitements.

### Facteurs non analysés

La fiche distribuée aux moniteurs comportait d'autres renseignements édaphiques et écologiques à fournir (description sommaire du sol par sa texture et sa couleur, présence d'autres plantes atteintes de malformations), ainsi que des renseignements agricoles complémentaires (rendement, date de buttage, traitements, récoltes). Trop de fiches fournissaient

ces données de façon rudimentaire ou nulle; nous n'avons pu les maintenir dans le fichier soumis à l'analyse quantitative.

D'autres facteurs, de nature qualitative (nature des produits insecticides, nature des antécédents culturels, par exemple), apparaîtront plus loin, après l'analyse par régression multiple ou dans l'analyse par segmentation.

## ÉTUDE DES VARIABLES DEUX À DEUX (projections croisées)

### Évolution en fonction du temps (campagnes)

Parmi l'ensemble des projections réalisées, nous ne retiendrons ci-dessous que celles relatives aux trois facteurs présentant des variabilités interannuelles.

#### Date de la première pluie

L'étalement des données relatives à cette date suivant les régions est résumé ci-après (découpage par 6 jours), pour chaque campagne :

Année	Date la plus précoce	Date la plus tardive
1970	78 <sup>e</sup> jour	177 <sup>e</sup> jour
1971	90 <sup>e</sup> "	202 <sup>e</sup> "
1972	59 <sup>e</sup> "	189 <sup>e</sup> "
1973	53 <sup>e</sup> "	208 <sup>e</sup> "
1974	109 <sup>e</sup> "	190 <sup>e</sup> "
1975	78 <sup>e</sup> "	190 <sup>e</sup> "
1976	84 <sup>e</sup> "	208 <sup>e</sup> "
1977	121 <sup>e</sup> "	208 <sup>e</sup> "
(1978)*	90 <sup>e</sup> "	196 <sup>e</sup> "

Étant donné la variabilité des contributions des régions dans l'enquête au fil des ans, ces indications sont assez sujettes à caution.

#### Date de semis

L'étalement observé, suivant les régions, est indiqué ci-après pour chaque campagne :

Année	Date la plus précoce	Date la plus tardive
1970	149 <sup>e</sup> jour	196 <sup>e</sup> jour
1971	140 <sup>e</sup> "	201 <sup>e</sup> "
1972	140 <sup>e</sup> "	183 <sup>e</sup> "
1973	138 <sup>e</sup> "	182 <sup>e</sup> "
1974	138 <sup>e</sup> "	192 <sup>e</sup> "
1975	133 <sup>e</sup> "	204 <sup>e**</sup> "
1976	142 <sup>e</sup> "	196 <sup>e</sup> "
1977	131 <sup>e</sup> "	183 <sup>e</sup> "
(1978)*	159 <sup>e</sup> "	192 <sup>e</sup> "

Les dates de semis sont donc beaucoup moins étalées que les dates de premières pluies.

### Nombre des applications insecticides

Voici, pour chaque année, les nombres d'applications relevés :

1970 : 1 à 5	1971 : 1 à 4	1972 : 0 à 5
1973 : 0 à 4	1974 : 0 à 6	1975 : 0 à 7
1976 : 0 à 5	1977 : 1 à 6	(1978)* : 0 à 5

D'autres projections « annuelles » fournies sur des facteurs agricoles par le calcul automatique, ne présentant qu'un intérêt très restreint, ne sont pas reprises dans cette étude.

### Autres projections croisées

Parmi les diverses figures des combinaisons obtenues, nous ne décrivons brièvement que celles présentant un intérêt général, en donnant pour chacune le coefficient de corrélation obtenu.

- Projection « nombre de pieds virescents recensés à l'hectare (= « NPVH ») et « surface du champ » : coefficient  $R = + 0,18$ .
- Projection « NPVH »  $\times$  « date de semis » ou « date de levée » :  $R = + 0,18$ .
- Projection « date de première pluie » et « date de semis » : elle présente un nuage avec un côté aplati (on sème rarement avant les pluies !); le coefficient de corrélation  $R = + 0,29$  traduit le fait banal que les semis sont effectués en fonction des pluies, mais de façon peu liée.
- Projection « date de première pluie » et « date de levée » : le nuage est encore plus aplati sur une face que pour la projection précédente (même si on a semé avant les pluies, la levée ne se fait pas avant celles-ci). Le coefficient  $R = + 0,23$  est du même ordre que précédemment.
- Projection « date de première pluie » et « durée entre pluie et levée » : elle montrerait que lorsque les premières pluies sont précoces, la germination est en général lente, et que celle-ci s'améliore en moyenne au milieu de saison. Le coefficient  $R = + 0,88$ . (N.B. : Si l'on remplace « levée » par « sarclage » dans le deuxième terme, le nuage est de même forme, incline de même façon, avec  $R = + 0,85$ .)

\* Données incomplètes pour 1978.

\*\* Donnée paraissant exceptionnelle.

## ANALYSE MULTIVARIÉE DES FACTEURS QUANTITATIFS

Une récapitulation des principaux résultats est présentée dans les tableaux suivants.

Le tableau 1 résume les données globales de surface et d'effectif échantillonnés par année et par ORD.

Les tableaux 2 et 3 indiquent la ventilation de la densité de l'attaque par zone (ORD ou secteur) et par année.

Les tableaux 4 à 9 opèrent la ventilation du taux d'attaque annuel moyen en fonction des principaux facteurs agricoles, pour les quatre ORD.

On procède à l'élaboration d'une régression multiple pas à pas, pour les facteurs quantitatifs; les fiches d'observations conservées pour cette étape portent sur 546 champs, pour lesquels les quinze variables figurant au tableau 10 étaient disponibles.

Tableau 1. — Surface et effectif échantillonnés par année et ORD

		Année									
		Ensemble	1970	1971	1972	1973	1974	1975	1976	1977	1978
ORD ensemble	Surface (ha)	1 226,63	173,02	306,22	38,82	91,98	86,50	181	236,52	98,66	13,91
	Effectif	823	120	185	37	75	58	112	143	64	29
Bobo .....	Surface (ha)	561,16	52,31	106,93	38,82	56,20	64,70	99,45	56,95	85,80	0
	Effectif	367	42	65	37	39	38	56	38	52	0
Dedougou .....	Surface (ha)	497,23	120,71	198,19	0	28,25	21,80	81,55	46,50	0,23	0
	Effectif	322	78	116	0	24	20	56	27	1	0
Diebougou .....	Surface (ha)	148,08	0	1,10	0	0	0	0	133,07	0	13,91
	Effectif	111	0	4	0	0	0	0	78	0	29
Ouagadougou ..	Surface (ha)	20,16	0	0	0	7,53	0	0	0	12,63	0
	Effectif	23	0	0	0	12	0	0	0	11	0

Tableau 2. — Nombre de plants viresscents/ha selon année et ORD

		Année									
		Ensemble	1970	1971	1972	1973	1974	1975	1976	1977	1978
ORD ensemble	Plants (ha)	159,06	208,38	125,86	278,35	104,31	184,16	193,09	157,55	99,52	113,52
	Ecart-type	269,50	297,96	190,89	400,09	202,00	301,31	382,84	238,50	151,43	119,42
	Nbre champs	823	120	185	37	75	58	112	143	64	29
Bobo .....	Plants (ha)	150,28	334,74	73,11	278,35	131,18	91,63	91,82	175,42	108,42	
	Ecart-type	255,40	380,32	93,92	400,09	265,81	161,76	182,48	196,95	164,43	
	Nbre champs	367	42	65	37	39	38	56	38	52	0
Dedougou .....	Plants (ha)	182,22	140,33	143,42		71,25	359,95	294,36	211,07	0	
	Ecart-type	298,97	213,24	213,56		88,92	407,91	489,21	194,95	0	
	Nbre champs	322	78	116	0	24	20	56	27	1	0
Diebougou .....	Plants (ha)	138,32		474,00					130,32		113,52
	Ecart-type	242,91		209,66					265,22		119,42
	Nbre champs	111	0	4	0	0	0	0	78	0	29
Ouagadougou ..	Plants (ha)	75,13				83,08				66,45	
	Ecart-type	63,33				67,73				56,90	
	Nbre champs	23	0	0	0	12	0	0	0	11	0



Tableau 3. — Nombre de plants viresscents/ha (par secteur et par campagne pour les seuls ORD de Bobo et Dédougou)

		Ensemble	Année								
			1970	1971	1972	1973	1974	1975	1976	1977	1978
Secteur ensemble	Plants (ha)	159,06	208,38	125,86	278,35	104,31	184,16	193,09	157,55	99,52	113,52
	Ecart-type	269,50	297,96	190,89	400,09	202,00	301,31	382,84	238,50	151,43	119,42
	Nbre champs	823	120	185	37	75	58	112	143	64	29
Bobo-Sud (Bobo) .....	Plants (ha)	216,04				459,17	265,71	42,22	255,83	150,64	
	Ecart-type	294,43				472,63	292,57	27,87	268,44	183,83	
	Nbre champs	45	0	0	0	6	7	9	12	11	0
N'Dorla (Bobo) .....	Plants (ha)	23,36		24,00		14,57		30,31	15,38		
	Ecart-type	26,68		22,28		13,05		30,74	15,77		
	Nbre champs	53	0	22	0	7	0	16	8	0	0
Houndé (Bobo) .....	Plants (ha)	233,45	334,74	98,53	338,94	14,00			192,94	68,38	
	Ecart-type	332,05	380,32	94,65	496,86	0			133,17	66,24	
	Nbre champs	114	42	19	18	1	0	0	18	16	0
Bobo-Nord (Bobo) .....	Plants (ha)	108,55		105,53		59,18	52,75	182,00		115,48	
	Ecart-type	177,76		121,76		77,29	71,99	264,71		190,72	
	Nbre champs	99	0	19	0	17	16	22	0	25	0
Orodara (Bobo) .....	Plants (ha)	122,05		69,40	220,95	154,88	51,87	30,33			
	Ecart-type	200,44		64,81	266,58	241,23	58,99	24,64			
	Nbre champs	56	0	5	19	8	15	9	0	0	0
Beromo (Dedougou) ..	Plants (ha)	309,81	321,27	270,85		71,25	422,94	566,76	240,74		
	Ecart-type	379,12	327,28	186,85		88,92	411,46	627,32	196,27		
	Nbre champs	144	22	33	0	24	17	25	23	0	0
Dedougou (Dedougou) ..	Plants (ha)	156,43	88,72	232,25				129,00	40,50		
	Ecart-type	340,66	49,77	329,68				61,90	29,71		
	Nbre champs	51	13	24	0	0	0	5	4	0	0
Nouna (Dedougou) ..	Plants (ha)	53,65	72,50	66,64			3,00	34,33			
	Ecart-type	61,24	42,93	70,95			2,94	48,55			
	Nbre champs	23	6	11	0	0	3	3	0	0	0
Kougny (Dedougou) ..	Plants (ha)	44,50	28,90	39,58				80,33			
	Ecart-type	64,18	18,87	40,83				116,44			
	Nbre champs	28	10	12	0	0	0	6	0	0	0
Solenzo (Dedougou) ..	Plants (ha)	59,04	90,21	34,33				63,82			
	Ecart-type	63,96	74,86	34,07				69,94			
	Nbre champs	52	14	21	0	0	0	17	0	0	0
Tougan (Dedougou) ..	Plants (ha)	16,77	24,86	13,00							
	Ecart-type	21,82	23,42	19,94							
	Nbre champs	22	7	15	0	0	0	0	0	0	0
Sourou (Dedougou) ..	Plants (ha)	60,00	120,00							0	
	Ecart-type	60,00	0							0	
	Nbre champs	2	1	0	0	0	0	0	0	1	0

Tableau 4. — *Nombre de plants virescents/ha selon ORD et nature du produit insecticide*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Produit insecticide ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	255,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Péprothion .....	Plants (ha)	167,41	114,40	332,08	80,92	
	Ecart-type	289,38	218,38	402,20	139,28	
	Nbre champs	216	118	59	39	0
Endrine-DDT-MP ..	Plants (ha)	193,52	250,58	223,78	140,65	71,44
	Ecart-type	354,66	416,62	402,41	213,16	56,74
	Nbre champs	106	24	49	17	16
Waly .....	Plants (ha)	149,82	111,31	151,91	216,30	
	Ecart-type	255,27	169,34	218,37	390,83	
	Nbre champs	120	49	44	27	0
Nuvacron/Azodrine	Plants (ha)	96,92	10,67		122,80	
	Ecart-type	117,69	8,58		122,81	
	Nbre champs	13	3	0	10	0
Endrine-DDT .....	Plants (ha)	39,85	37,45	44,24		
	Ecart-type	49,17	52,85	41,28		
	Nbre champs	48	31	17	0	0
Mélanges divers ...	Plants (ha)	180,61	164,09	275,10	147,86	10,00
	Ecart-type	256,82	215,17	397,44	131,65	0
	Nbre champs	51	33	10	7	1
« Inconnu » .....	Plants (ha)	163,16	217,76	125,38	154,82	95,83
	Ecart-type	243,48	299,69	192,05	172,07	74,95
	Nbre champs	268	108	143	11	6

Tableau 5. — *Nombre de plants virescents/ha par ORD, « avec » ou « sans » traitement*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Protection insecticide ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	255,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Non traité .....	Plants (ha)	100,11	112,80	35,00	27,00	240,00
	Ecart-type	106,78	116,97	0	0	0
	Nbre champs	9	5	2	1	1
Traité .....	Plants (ha)	159,72	150,80	183,14	139,33	67,64
	Ecart-type	270,68	256,76	299,68	243,78	53,86
	Nbre champs	814	362	320	110	22

Tableau 6. — *Nombre de plants virescents/ha selon ORD et nombre d'applications*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Nombre de traitements ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	255,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Non traité .....	Plants (ha)	100,11	112,80	35,00	27,00	240,00
	Ecart-type	106,78	116,97	0	0	0
	Nbre champs	9	5	2	1	1
1 à 2 traitements .	Plants (ha)	125,90	114,53	157,83	97,50	57,42
	Ecart-type	244,85	210,26	304,23	156,30	46,20
	Nbre champs	198	49	92	38	19
3 à 4 traitements .	Plants (ha)	166,16	147,43	197,44	145,45	132,33
	Ecart-type	274,45	254,52	301,47	260,45	53,94
	Nbre champs	573	290	218	62	3
5 à 6 traitements .	Plants (ha)	233,52	280,05	104,40	260,30	
	Ecart-type	313,28	334,93	156,41	344,67	
	Nbre champs	42	22	10	10	0
7 et + .....	Plants (ha)	61,00	61,00			
	Ecart-type	0	0			
	Nbre champs	1	1	0	0	0

Tableau 7. — *Nombre de plants virescents/ha selon ORD et « avec » ou « sans » fertilisation*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Fertilisation ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	255,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Sans engrais .....	Plants (ha)	115,91	136,94	84,85	164,50	74,93
	Ecart-type	213,63	235,62	105,96	337,96	60,58
	Nbre champs	156	50	61	30	15
Avec engrais .....	Plants (ha)	169,39	152,39	204,93	128,62	83,86
	Ecart-type	280,16	258,32	323,91	195,47	68,95
	Nbre champs	666	317	261	81	7



Tableau 8. — *Nombre de plants viresscents/ha par ORD et précédent cultural*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Précédent cultural ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	253,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Jachère .....	Plants (ha)	123,00	132,25	151,83	91,28	17,00
	Ecart-type	201,38	235,49	210,17	119,49	7,00
	Nbre champs	123	56	29	36	2
Arachide .....	Plants (ha)	176,49	164,20	220,69	162,38	84,00
	Ecart-type	250,37	207,40	304,53	257,96	47,10
	Nbre champs	39	15	13	8	3
Coton .....	Plants (ha)	157,60	169,17	147,73	196,70	27,67
	Ecart-type	253,32	264,88	249,29	225,46	15,35
	Nbre champs	159	84	59	10	6
Céréales .....	Plants (ha)	169,82	135,46	206,64	159,92	108,40
	Ecart-type	283,97	210,87	333,16	306,25	68,57
	Nbre champs	411	166	185	50	10
Tubercules .....	Plants (ha)	161,17	161,17			
	Ecart-type	170,32	170,32			
	Nbre champs	6	6	0	0	0
Sésame .....	Plants (ha)	94,35	107,33	34,00	102,00	
	Ecart-type	254,47	285,70	24,63	0	
	Nbre champs	23	18	4	1	0
Inconnu .....	Plants (ha)	175,90	258,59	135,06	117,17	96,00
	Ecart-type	335,99	471,49	230,94	163,27	48,00
	Nbre champs	62	22	32	6	2

Tableau 9. — *Nombre de plants viresscents/ha selon ORD et « avec » ou « sans » buttage*

		ORD				
		Ensemble	Bobo	Dedougou	Die-bougou	Ouaga-dougou
Buttage ensemble	Plants (ha)	159,06	150,28	182,22	138,32	75,13
	Ecart-type	269,50	253,40	298,97	242,91	63,33
	Nbre champs	823	367	322	111	23
Non butté .....	Plants (ha)	141,74	163,43	140,56	88,26	73,71
	Ecart-type	270,41	299,87	274,08	143,79	76,60
	Nbre champs	302	132	121	42	7
Butté .....	Plants (ha)	169,11	142,90	207,30	168,78	75,75
	Ecart-type	268,46	226,30	310,32	282,64	56,54
	Nbre champs	521	235	201	69	16

Tableau 10. — *Liste des variables quantitatives utilisées dans l'analyse en régression multiple*

N° de la variable	Enoncé de la variable	Terme corresp.
1	$\sqrt{\text{nombre plants viresscents/ha} \times 100}$ .....	Y
2	$\sqrt{\text{surface du champ en ares} \times 100}$ .....	X <sub>1</sub>
3	Date levée (en jours calendaires) .....	X <sub>2</sub>
4	Date semis (en jours calendaires) .....	X <sub>3</sub>
5	Date 1 <sup>er</sup> sarclage (en jours calendaires) .....	X <sub>4</sub>
6	Nombre de traitements .....	X <sub>5</sub>
7	Date 1 <sup>er</sup> traitement (en jours calendaires) .....	X <sub>6</sub>
8	Date 1 <sup>re</sup> pluie (en jours calendaires) .....	X <sub>7</sub>
9	Durée « levée - 1 <sup>er</sup> traitement » (en jours) .....	X <sub>8</sub>
10	Durée « 1 <sup>re</sup> pluie - levée » (en jours) .....	X <sub>9</sub>
11	Durée « 1 <sup>re</sup> pluie - sarclage » (en jours) .....	X <sub>10</sub>
12	Durée « semis - 1 <sup>er</sup> traitement » (en jours) .....	X <sub>11</sub>
13	Durée « semis - 1 <sup>er</sup> sarclage » (en jours) .....	X <sub>12</sub>
14	Durée « 1 <sup>re</sup> pluie - 1 <sup>er</sup> traitement » (en jours) .....	X <sub>13</sub>
15	Nombre de sarclages .....	X <sub>14</sub>

Le traitement automatique des données suivant le programme d'analyse « Multivar » fut effectué au

CITIM\*\*\*, sur IBM 360. Les principaux résultats sont résumés dans les tableaux 11, 12 et 13.

Tableau 11. — *Régression multiple pas à pas : classement des variables d'après leur corrélation décroissante avec la variable à expliquer*

N°	Libellé des variables	Moyenne	Ecart-type	Coefficient de corrélation
1	$\sqrt{\text{RPV/ha} \times 100}$ .....	100,0385	81,6802	1,0000
2	$\sqrt{\text{Surf.} \times 100}$ .....	114,9084	38,7462	— 0,2144
3	Date levée .....	173,1099	14,2768	— 0,1873
4	Date semis .....	162,8883	12,7312	— 0,1779
5	Date 1 <sup>er</sup> sarclage .....	190,7106	16,4263	— 0,1507
6	Nombre de traitements .....	3,1740	0,9420	0,1487
7	Date 1 <sup>er</sup> traitement .....	219,3791	18,9095	— 0,1002
8	Date 1 <sup>re</sup> pluie .....	142,1813	28,9488	— 0,0515
9	Levée - date de 1 <sup>er</sup> traitement .....	46,2692	18,7275	0,0416
10	Pluie - levée .....	30,9286	29,1812	— 0,0406
11	Pluie - sarclage .....	48,5293	30,3646	— 0,0324
12	Semis - 1 <sup>er</sup> traitement .....	56,4908	17,6800	0,0209
13	Semis - sarclage .....	27,8223	13,9080	— 0,0152
14	Pluie - 1 <sup>er</sup> traitement .....	77,1978	32,2657	— 0,0125
15	Nombre de sarclages .....	1,7344	0,4775	0,0091

Dès l'abord, on considère une formule réduite à sept termes explicatifs car, au-delà, les variables ne sont plus prises en compte (parce que trop corrélées avec d'autres, ou apportant moins de 1 % d'explication de variance [tabl. 12]).

Dans cette formule, « date de semis » et « date de levée » sont fortement corrélées ; on peut éliminer l'un ou l'autre facteur sans perdre de précision. Les

variables « date de première pluie » et « date de sarclage » sont aussi plus ou moins liées aux précédentes, et en tous cas de très faible signification, de même que « date du premier traitement ».

On arrive donc à une formule ne comportant plus que trois termes, mais où chacun de ces termes manifeste un test « F » de Snedecor de signification élevée (tabl. 13). Malgré cela, le « pourcentage total expliqué » reste faible (10,94 % contre 11,41 % avec sept termes), ce qui montre clairement que des facteurs décisifs ne sont pas tabulés dans les questionnaires remis lors de l'enquête.

\*\*\* CITIM : Centre interuniversitaire de traitement de l'information de Montpellier.

Tableau 12. — Régression multiple pas à pas :  
tableau des coefficients et critères pour sept variables explicatives

Libellé	Coefficient du polynome	Coefficient réduit	Erreur standard relative sur le coefficient	Variance marginale (en %)	F de Snedecor (538 d.d.l.)
$\sqrt{\text{Surf.}} \times 100$ .....	— 0,5427D 00	— 0,2574	0,16	6,33	38,47
Date levée .....	— 0,5470D 00	— 0,0956	0,70	0,33	2,03
Date semis .....	— 0,6561D 00	— 0,1023	0,69	0,35	2,12
Date 1 <sup>er</sup> sarclage ....	— 0,1137D 00	— 0,0229	2,24	0,03	0,20
Nbre de traitements.	0,1075D 02	0,1240	0,35	1,38	3,40
Date 1 <sup>er</sup> traitement ..	— 0,6943D-01	— 0,0161	2,91	0,02	0,12
Date 1 <sup>re</sup> pluie .....	0,3606D-01	0,0128	3,33	0,01	0,09
Variance expliquée = 761,54.					
Pourcentage expliqué = 11,41.					
Ecart-type résiduel = 76,88.					
Terme constant = 361,621115.					

Tableau 13. — Régression multiple pas à pas :  
formule condensée à trois termes explicatifs

Libellé	Coefficient du polynome	Coefficient réduit	Erreur standard relative sur le coefficient	Variance marginale (en %)	F de Snedecor (542 d.d.l.)	Contrib. relative (% variance expliq.)
$\sqrt{\text{Surf.}} \times 100$ ...	— 0,5434D 00	— 0,2578	0,16	6,45	39,24 **	4,60
Date semis .....	— 0,1227D 01	— 0,1913	0,22	3,44	20,91 **	4,60
Nb. traitements .	0,1170D 02	0,1349	0,31	1,74	10,57 **	1,74
						T. 10,94
Variance expliquée = 729,61.						
Pourcentage expliqué = 10,94 %.						
Ecart-type résiduel = 77,08.						
Terme constant = 325,24.						

#### Formule proposée

Avec  $Y = \sqrt{\text{nombre de pieds viresscents/ha}} \times 100$ ,

$X_1 = \sqrt{\text{surface du champ}} \times 100$ ,

$X_2 = \text{date de semis}$ ,

$X_3 = \text{nombre de traitements}$ ,

on a la relation suivante :

$$Y = 325,24 - 0,54 X_1 - 1,23 X_2 + 11,7 X_3$$

dans laquelle l'accroissement de la surface du champ et le retard de la date de semis correspondent à une légère diminution du nombre de pieds viresscents, et où un nombre de traitements accru correspond à un accroissement du taux de la maladie.

## ÉTUDE DES FACTEURS QUALITATIFS

Ayant poussé aussi loin que possible l'analyse statistique en régression multiple des facteurs quantifiables, il reste à étudier les autres facteurs mentionnés dans les fiches, non exploitables quantitativement. Cette étude peut porter sur des fichiers d'amplitude variable.

### Nature des produits insecticides (tabl. 4)

L'étude a porté sur le fichier complet (823 fiches) ; on a dû créer des catégories telles que « mélanges » (ou succession) d'insecticides, ou encore produit « inconnu » (non mentionné dans les fiches, plus

exactement). La participation inégale des divers ORD est un inconvénient certain qui est retrouvé ici.

Le classement moyen d'ensemble qui apparaît est le suivant :

Matieres actives	Taux de maladie
Endrine-DDT .....	46 plants/ha
Monocrotophos-DDT .....	97 " "
PCC-DDT-MP .....	150 " "
« Inconnu » .....	163 " "
Péprothion .....	168 " "
Mélanges .....	180 " "
Endrine-DDT-MP .....	193 " "

En fait, l'écart-type étant dans tous les cas supérieur à la valeur moyenne, on ne peut s'autoriser à tirer de conclusion bien assurée des chiffres du tableau. On peut ainsi noter qu'endrine-DDT, suivant que le mélange est complété ou non par le méthyl-parathion, se trouve soit en queue, soit en tête du classement ; celui-ci est donc illusoire. En plus, l'effet « année » (qui n'apparaît pas séparément ici) est important, car les produits appliqués ont varié au cours de la période envisagée.

#### Effets des traitements (tabl. 5 et 6)

On a vu que la formule de régression simplifiée englobait l'effet quantitatif « nombre de traitements » comme terme en troisième position, avec une participation faible, mais négative. Ici, l'opposition qualitative entre « non traité » et « traité » indique que l'on passerait de 100 à 160 plants malades/ha ; mais la faible valeur de la catégorie « non traité » (avec 9 champs seulement) et l'écart-type supérieur à

la moyenne correspondante incitent à ne tirer qu'une conclusion prudente.

#### Effets des précédents cultureux (tabl. 8)

En résumé, on obtient le classement suivant :

Précédent cultural	Taux de maladie
Arachide .....	176
Inconnu .....	176
Céréales .....	179
Tubercules .....	161
Cotonnier .....	158
Jachère .....	123
Sésame .....	94

Le précédent « cotonnier » ou « sésame » ne paraît pas spécialement favorable à l'implantation de la maladie sur le même champ cultivé en cotonnier à la campagne suivante ; par ailleurs, aucune conclusion statistique n'est possible, en raison des écarts-types élevés par rapport à la moyenne correspondante.

#### Effets de la fumure et du buttage

Les tableaux 7 et 9 ne révèlent aucune différence nette entre les champs fumés ou non, ni entre champs buttés ou non ; de plus, les écarts-types sont toujours élevés.

#### Conclusion sur l'étude des facteurs qualitatifs

On constate que les contributions venant des facteurs qualitatifs sont confuses et imprécises, probablement du fait des biais introduits par l'échantillonnage.

## DISCUSSION SUR L'ANALYSE DES FACTEURS QUANTITATIFS ET QUALITATIFS

### Relation avec la dimension des champs

Ce facteur, qui ressort en premier lieu de l'analyse en régression multiple, est affecté d'un signe négatif ; comme on a transformé la surface par « racine carrée », on peut dire que l'incidence de la maladie est d'autant plus faible qu'une « dimension linéaire » du champ est grande.

L'observation courante montre que, dans de nombreux cas, ce sont les bordures exposées à la brousse qui sont les plus atteintes par la maladie ; derrière ces zones à taux de contamination plus élevé, la fréquence de l'infection tend à s'abaisser, bien qu'irrégulièrement, avec la profondeur du champ.

La recommandation de regrouper les champs, dans la mesure où la surface utilement exploitable le permet, est donc valable tant pour les parasites directs en général (phytophages, chenilles des organes fructifères, punaises de capsules) que pour la virescence florale, en fonction vraisemblablement de l'origine externe des migrations de l'insecte-vecteur

et des déplacements de celui-ci, orientés vers les champs de cottonniers.

### Relation avec la date de semis

Ce facteur apparaît en second, par ordre d'importance, dans la régression multiple ; il est doté d'un signe négatif : plus la date de semis est tardive, plus le taux moyen de maladie s'abaisse. Cette conclusion rejoint les résultats de l'expérimentation conduite depuis plusieurs années sur la ferme de multiplication de Boni. Cette concordance de résultats permet de généraliser à la culture paysanne les conclusions obtenues, en conditions expérimentales, à partir de parcelles de faible étendue.

### Relation avec le nombre des applications foliaires d'insecticides

Les techniques de protection par imprégnation de semences ou par application au sol de produits insecticides spéciaux ne sont pas passées en vulgarisation,

malgré qu'elles se soient montrées très efficaces en expérimentation, comme le relaient les rapports annuels de Haute-Volta depuis plusieurs années.

La présente enquête, réalisée chez les paysans, n'envisage que les applications foliaires d'insecticides conventionnels, applications qui débutent généralement entre le 50<sup>e</sup> et le 60<sup>e</sup> jour après le semis ; elles ne peuvent de façon sensible réduire le taux d'incidence de la maladie, compte tenu des observations suivantes :

1<sup>o</sup> La contamination se fait essentiellement lors des stades jeunes de la plante, donc bien avant la mise en œuvre des traitements foliaires, qui ne débutent pas avant le 50<sup>e</sup> jour dans la meilleure hypothèse.

2<sup>o</sup> Le délai d'apparition des symptômes étant très long (au moins 30 à 40 jours après l'infection), même si les traitements foliaires avaient une certaine incidence favorable sur le taux de la contamination par l'insecte, leur effet quant aux symptômes apparaîtrait trop tardivement en cours de végétation et n'influerait même pas sur les premières récoltes.

Il reste toutefois à présenter des hypothèses cohérentes sur la relation à signe positif mise en avant ici : « le taux de maladie serait plus élevé dans les champs les plus traités ». Il ne peut s'agir à l'évidence d'une liaison directe de cause à effet. Nous suggérons, pour expliquer ce paradoxe apparent, l'existence d'une relation indirecte entre la date de semis, le développement végétatif des plantes et la date d'infection ; on sait, en effet, que les semis les plus précoces sont en moyenne les plus attaqués par la virescence florale, comme le montrent les essais en Haute-Volta, probablement parce qu'il y a meilleure coïncidence entre le stade sensible de la plante et l'époque où les insectes sont des contaminants plus efficaces. Les migrations des populations, variables en intensité, peuvent se produire à diverses époques, notamment en mai-juin, puis plus tard, en septembre ; mais le facteur qui semble jouer le rôle décisif n'est pas tant le nombre d'individus migrant à telle date, que le taux d'individus infectants présents dans cette population ; des cultures réalisées sur milieux spéciaux pour mycoplasmes (GIANNOTTI et DELATTRE, 1973) ont montré que le taux d'individus porteurs de mycoplasmes décroît généralement rapidement avec le temps, à partir de fin mai. Les chances de contamination des jeunes plants s'atténuent donc en général assez rapidement, et ceci pour une double raison. Or, les cultures ayant bénéficié de semis précoces reçoivent naturellement un nombre moyen de traitements plus élevé, leur période de végétation active étant au total plus longue, et les champs ayant un potentiel plus élevé reçoivent aussi plus de soins de la part des cultivateurs que les champs semés plus tardivement.

La relation tout à fait inattendue entre précocité de semis et nombre de traitements peut donc s'établir, indirectement, à travers ces divers facteurs, taux de maladie et potentiel de production particulièrement élevé.

### Relation avec le facteur « précédents culturaux »

L'absence d'indication nette d'un effet néfaste d'un précédent cotonnier ou sésame tend à renforcer l'idée que ni l'une ni l'autre de ces plantes n'est un réservoir d'inoculum important pour la maladie ou une plante-hôte servant de pont pour l'insecte-vecteur pendant l'intersaison. Il importe donc de poursuivre les investigations pour rechercher les plantes sauvages qui jouent éventuellement l'un ou l'autre rôle, source d'inoculum de la maladie et plante-hôte du vecteur.

### Relation avec le rendement

Cette information constituait l'une des motivations primaires de l'enquête ; malheureusement, le facteur « rendement » n'a pu être mis correctement en œuvre dans nos fichiers d'analyses, pour la raison pratique qu'assez peu de fiches comportaient des renseignements à ce sujet et que, parmi celles qui faisaient état de rendement, la plupart étaient visiblement très peu fiables.

Nous pouvons seulement rappeler ici deux faits incontestables obtenus tant par l'observation courante que par l'expérimentation classique :

1<sup>o</sup> Pour chaque plant atteint, la production de capsules cesse complètement dès l'apparition des premiers symptômes ; la perte est donc mathématiquement proportionnelle au taux de contamination quand celle-ci est précoce ; elle est liée, d'autre part, au délai qui reste à courir depuis la manifestation des symptômes jusqu'à l'époque de fin de floraison utile, lorsque l'attaque a lieu au cours de la floraison.

2<sup>o</sup> Les champs les plus productifs sont, en général, ceux où le semis et la levée sont les plus précoces ; ils sont, par là-même, souvent pénalisés par un taux de maladie plus élevé, comme l'étude en régression multiple l'indique et comme les observations courantes le montrent ; mais cette pénalisation, sauf exceptions rarissimes, reste bien inférieure au gain de potentiel de production résultant directement d'un semis précoce ; d'où le résultat paradoxal : « les champs les plus productifs en raison de leur précocité sont souvent, et pour la même raison, les plus contaminés », ce qui ne doit quand même pas être exprimé sous une forme laissant entendre que la « virescence accroît la productivité des plants » ! De même, en pratique, nous ne pouvons pas souscrire aux recommandations de retarder les semis, faites par d'autres auteurs (DESMIDIS, 1975).



## ANALYSE GLOBALE PAR LA MÉTHODE DE SEGMENTATION

## Principe

On appelle « segmentation » une méthode de classification établissant par dichotomies successives des groupes de champs à la fois homogènes vis-à-vis de la variable à expliquer (ici : « nombre de plants viresscents ») et les plus différents possibles entre eux, par découpage du fichier selon les modalités des variables les plus explicatives. Cette méthode permet, en outre, une approche explicative voisine de celle fournie par la régression : en effet, on mesure à chaque itération, l'explication de la variance appliquée par la classification et ainsi, indirectement, on hiérarchise l'importance relative des variables explicatives utilisées pour l'élaboration des groupes. La segmentation a, de plus, l'avantage de pouvoir traiter des problèmes de nature complexe, par exemple des mélanges de variables quantitatives et qualitatives.

A partir de la population totale, et par itérations successives, le programme recherche les dichotomies selon les variables explicatives qui maximisent la variance expliquée de la variable « nombre de plants viresscents ».

A la fin de chaque itération, une analyse de variance à un critère de classification permet de vérifier, par un test F, s'il existe encore des différences significatives entre les groupes formés. Ceci permet de fixer le niveau de partition au-delà duquel un nouveau découpage n'aurait plus de sens statistique. De plus, l'expérimentateur fixe un seuil minimal d'individus par segment, et un nombre maximal de segments, pour rester dans les limites du bon sens.

On aboutit finalement à un arbre de classification où chaque ramification s'établit clairement par une dichotomie « optimale ».

## Données de base

Les 13 variables utilisées pour cette analyse ont été les suivantes :

1.  $\sqrt{\text{nombre de plants viresscents à l'ha} \times 100}$  = variable à expliquer quantitative ;
2. année = variable explicative qualitative à 9 éventualités ;
3. ORD = variable explicative qualitative à 4 éventualités ;
4. précédent cultural = variable explicative qualitative à 6 éventualités ;
5. engrais = variable explicative qualitative ordonnée à 2 éventualités ;
6. traitements = variable explicative qualitative ordonnée à 5 éventualités ;
7. sarclages = variable explicative qualitative ordonnée à 3 éventualités ;
8. époque première pluie = variable explicative qualitative ordonnée à 5 éventualités ;

9. époque semis = variable explicative qualitative ordonnée à 5 éventualités ;
10. époque levée = variable explicative qualitative ordonnée à 5 éventualités ;
11. époque premier sarclage = variable explicative ordonnée à 3 éventualités ;
12. époque premier traitement = variable explicative ordonnée à 5 éventualités ;
13. surface du champ = variable explicative qualitative ordonnée à 5 éventualités.

Le nombre de plants viresscents a été conservé en données quantitatives, mais les variables explicatives ne peuvent être que qualitatives. Certaines l'étaient au départ (année, ORD, précédent cultural), les autres ont été transformées par découpage en classes au vu des histogrammes. De plus, l'ordre des éventualités n'étant pas quelconque, elles ont été définies en qualitatives ordonnées.

Le fichier sans donnée manquante sur l'une ou l'autre des variables, après filtrage de la population totale, aboutit à un sous-fichier de 494 individus complets quant à ces 13 variables, soit un échantillonnage un peu différent de celui de la régression multiple.

Les options prises pour arrêter les calculs ont été les suivantes :

- nombre maximal de segments = 32 ;
- effectif minimal pour qu'un groupe puisse être segmenté = 20 individus ;
- effectif minimal des sous-groupes pour qu'une segmentation puisse être acceptée = 10 individus.

## Processus d'établissement de la segmentation

Dans le cas présent, on part d'une population de 494 champs, sur lesquels la moyenne globale du nombre de plants viresscents à l'hectare s'établit à 99  $[(99,5)/100]$ .

*Première itération :* la meilleure dichotomie est obtenue avec l'époque de levée ; la population se répartit ainsi en deux groupes significativement différents entre eux :

- 1) Levée jusqu'au 175<sup>e</sup> jour : 283 champs, avec une moyenne de 135 plants viresscents  $[(116,3)/100]$ .
- 2) Levée après le 175<sup>e</sup> jour : 211 champs, avec une moyenne de 59 plants viresscents à l'hectare.

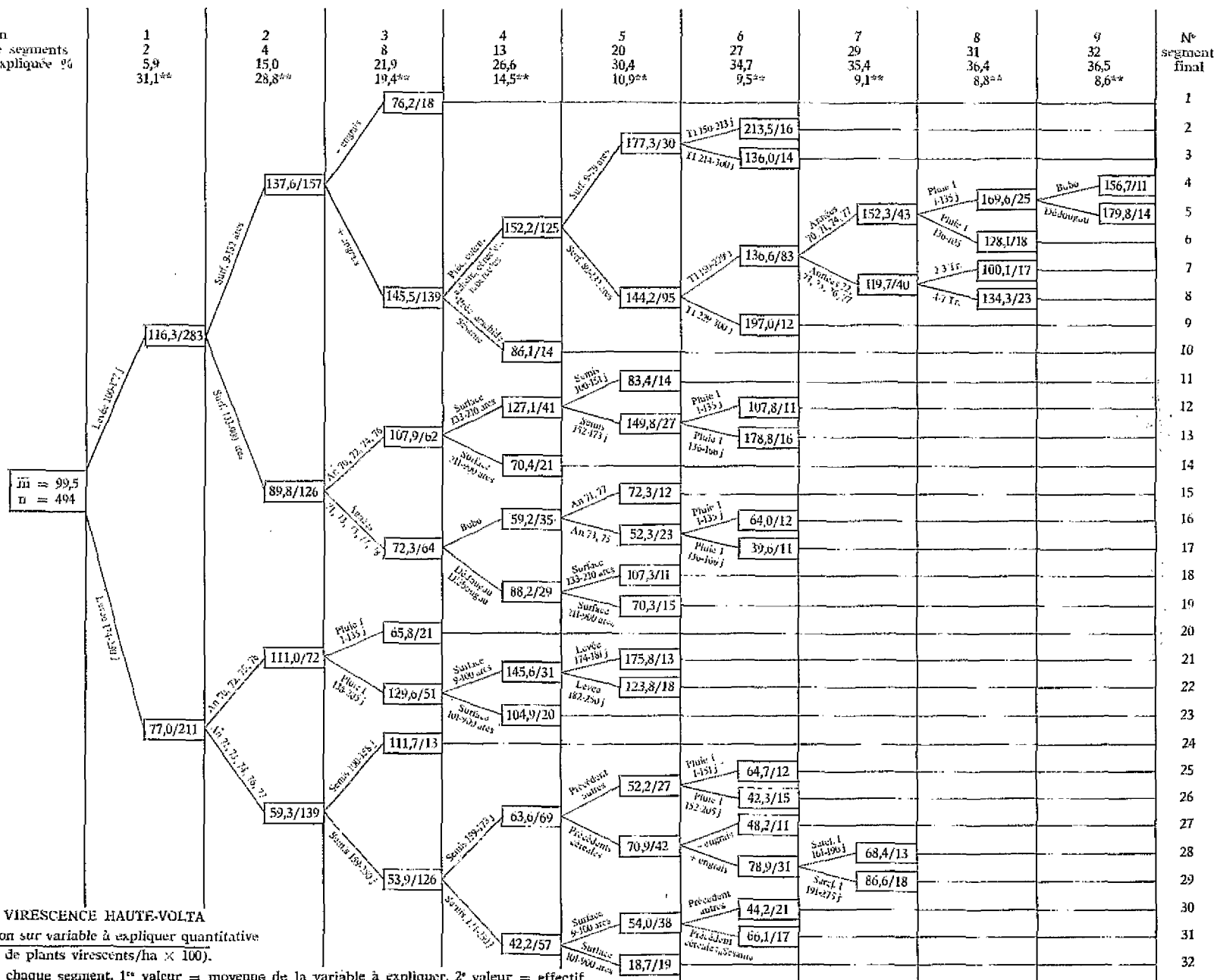
Ce découpage apporte une contribution de 6 % à l'explication de la variance.

*Seconde itération :* le groupe supérieur peut être lui-même découpé en 2 sous-segments, en fonction de la surface des champs :

- 1) Surface inférieure à 133 ares : 157 champs, avec moyenne = 188,2 pl. vir.
- 2) Surface supérieure à 133 ares : 126 champs, avec moyenne = 30,6 pl. vir.



Arbre dichotomique et ses principales caractéristiques.



tandis que le deuxième groupe peut être dichotomisé en fonction de l'année :

3) Années 1970, 1972, 1975, 1978 : 72 champs, avec moyenne = 123,2 pl. vir.

4) Années 1971, 1973, 1974, 1976, 1977 : 139 champs, avec moyenne = 35,2 pl. vir.

L'explication de la variance est alors portée à 15 % par ce découpage en 4 groupes, les différences entre groupes restant hautement significatives ( $F = 28,8$ ).

*Troisième itération* : chacun des 4 groupes précédents est scindé en deux, ce qui donne 8 segments.

Le premier groupe (levée précoce, petite surface) est découpé en fonction de la présence ou non d'engrais :

1) Sans engrais : 18 champs, moyenne = 58,1 pl. vir.

2) Avec engrais : 139 champs, moyenne = 211,7 pl. vir.

Le deuxième groupe (levée précoce, grande surface) est découpé en fonction de l'année :

3) Années 1970, 1972, 1974, 1976 : 62 champs, avec moyenne = 116,4 pl. vir.

4) Années 1971, 1973, 1975, 1977, 1978 : 64 champs, avec moyenne = 52,3 pl. vir.

Le troisième groupe (levée tardive, années 1970, 1972, 1975, 1978) est découpé en fonction de l'époque de la première pluie :

5) Pluie précoce (avant le 135<sup>e</sup> jour) : 21 champs, avec moyenne = 43,3 pl. vir.

6) Pluie tardive (du 136<sup>e</sup> au 205<sup>e</sup> jour) : 51 champs, avec moyenne = 170,0.

Le quatrième groupe (levée tardive, années 1971, 1973, 1974, 1976, 1977) est découpé selon l'époque de semis :

7) Semis précoces (avant le 159<sup>e</sup> jour) : 13 champs, avec moyenne = 124,8 pl. vir.

8) Semis moyens et tardifs (du 159<sup>e</sup> au 250<sup>e</sup> jour) : 126 champs, avec moyenne = 34,7 pl. vir.

A ce stade, on atteint 22 % d'explication de variance, avec des différences encore hautement significatives entre certains des nouveaux groupes ( $F = 19,4$ ).

Le lecteur pourra poursuivre le processus d'analyse sur l'arbre de classification donné page 181.

jusqu'à la branche finale non dichotomisée, appelée « segment final ».

## Discussion

L'analyse par segmentation donne l'époque de levée des cotonniers comme le facteur le plus discriminant : les champs à levée précoce sont nettement plus contaminés.

Dans le groupe des levées précoces, la surface du champ intervient ensuite et de façon négative : puis, pour les petites surfaces, l'apport d'engrais se montre comme un facteur défavorable, tandis que, sur les grandes surfaces, on enregistre d'abord un effet de l'année (1970, 1972, 1974 et 1976 étant plus contaminées), puis un effet « surface » sur les catégories des champs les plus atteints et, enfin, un effet régional sur les champs les moins atteints.

Sur les levées tardives, on observe à nouveau un effet variable suivant l'année. Dans les années favorables, la tardiveté des semis intervient ensuite à plusieurs reprises pour affiner progressivement la classification des champs les moins atteints (avec « R.P.V.H. » toujours < 100 dans ces groupes). Lors des années les moins favorables, on observe un effet de la « tardiveté de la première pluie », suivi d'un effet « surface » et à nouveau d'un effet « tardiveté de la levée ».

A la cinquième itération, on parvient ainsi à expliquer 30 % de la variance du nombre de plants virecents, avec une classification en 14 groupes, intégrant à la fois des facteurs quantitatifs et qualitatifs. Au-delà, des nuances supplémentaires peuvent être apportées dans les groupes plus volumineux ; en poussant l'exploitation à l'extrême, une classification en 32 segments apporte 36,5 % d'explication de variance (à la neuvième itération).

Certains groupes arrivent à des moyennes identiques par des voies dichotomiques très différentes ; à ceci correspond en partie une approche explicative des premiers « pas » moins satisfaisante en régression multiple, qui fournit l'analyse globale en une seule fois et de façon linéaire. Il y aurait donc intérêt à utiliser une classification dichotomique en premier lieu, pour établir ensuite des formules de régression multiple propres aux différentes partitions pratiquées dans le fichier de base (par exemple, à partir des 4 segments de l'itération 2).

## DISCUSSION GÉNÉRALE

### Confrontation des résultats obtenus par les différentes méthodes

L'effet de la surface du champ apparaît en première position en régression multiple, mais il n'explique globalement que 4,6 % de la variance. La segmentation montre que cet effet doit être modulé : il n'intervient fortement (2<sup>e</sup> itération, puis 4<sup>e</sup> et 5<sup>e</sup>) que sur des champs levés précocement ; sur les champs tardifs, il n'intervient qu'en 4<sup>e</sup> et 5<sup>e</sup> itérations pour introduire des nuances supplémentaires. Cet

effet de surface est résumé au tableau 14, où l'on notera l'excellent ordonnancement obtenu au plan quantitatif ; ce qui explique son rôle en régression multiple.

L'effet « date des premières pluies », retenu dans le polynôme à 7 variables, n'est pas repris pour le polynôme à 3 variables ; de même, les effets « date de semis » et « date de levée » ont été confondus dans la formule simplifiée de régression multiple, parce qu'assez fortement corrélés. La segmentation

Tableau 14. — *Segmentation : effet de la surface du champ sur le taux d'infestation par la virescence*

	Surface (ares)		Nbre plants virescents/ha $(\sqrt{x \times 100})$
1. Levée précoce (100-173 <sup>e</sup> jour)			
avec engrais et précédents coton, céréales, tubercules ou jachère .....	\ 9-79	→	177,3
	/ 80-132	→	144,2
années 70-72-74-76 .....	\ 133-210	→	127,1
	/ 211-900	→	70,4
années 71-73-75-77-78 et ORD de Diébougou et Dédougou ..	\ 134-210	→	107,3
	/ 211-900	→	70,3
2. Levée tardive (174-250 <sup>e</sup> jour)			
années 70-72-75-78 et 1 <sup>re</sup> pluie tardive (136-205 <sup>e</sup> jour) .....	\ 9-100	→	145,6
	/ 101-900	→	104,9
années 71-73-74-76-77 et semis tardifs (174-250 <sup>e</sup> jour) .....	\ 9-100	→	54,0
	/ 101-900	→	18,7

permet de distinguer ces facteurs d'après les différents niveaux d'action. La « date des premières pluies » apparaît en général dans les itérations tardives (soit les ultimes, soit les pénultièmes), ou bien tout simplement elle ne figure même pas au tableau (segments n° 1-2-3-7-8-9-10-11-14-15-18-19-27-28-29-30-31-32) ; seules exceptions, les segments n° 20-21-22-23 (antépénultièmes). Il semble donc confirmé que ce facteur soit peu décisif, dans l'un et l'autre processus d'analyse.

Quand la « date de semis » apparaît, elle précède le facteur « date de première pluie » si ce dernier intervient ; mais elle peut aussi ne pas se présenter, alors que « date de première pluie » figure parfois dans la dichotomie (par exemple n° 20 à 23).

Le facteur « date de levée », qui joue dès la première itération, n'apparaît plus ensuite qu'une seule fois et en fin de sectionnement, dans la cinquième itération (segments n° 21 et 22) ; il donne alors une indication peu conforme à la première impression, puisque les taux de maladie sont très élevés (175,8 et 123,8) et bien supérieurs aux moyennes données dans la première dichotomie (77,0 et 116,3). Or, cette importante correction a lieu après élimination des facteurs « première pluie : précoce » et « surface de champ : importante ». On peut en déduire que certaines années (1970, 1972, 1975, 1978), la règle « taux de virescence d'autant plus élevé que la levée est précoce » n'est pas rigoureusement suivie, et émettre l'hypothèse que les petites surfaces, lorsque les premières pluies sont tardives — avec une levée  $\pm$  tardive, par conséquent, mais aussi probablement sortie des adultes des insectes-vecteurs retardée — peuvent alors être infestées bien au-dessus du taux moyen attendu.

D'une façon générale, la segmentation permet donc d'appréhender, de façon beaucoup plus fine qu'une régression multiple brutale, un problème complexe

où de nombreuses interactions se manifestent entre les variables. Alors que la régression permet à peine de dépasser 10 % d'explication de variance, la segmentation permet d'atteindre 30 % par une simple stratification du fichier en 14 groupes, à l'intérieur desquels la hiérarchie des facteurs explicatifs peut varier très sensiblement.

### Intérêt et défauts de ce type d'enquête

L'intérêt primordial d'une enquête phytosanitaire est de pouvoir relier ses résultats à ceux tirés de l'expérimentation classique et de justifier le bien-fondé de l'extrapolation de ces derniers.

Certains facteurs, clairement analysables dans les protocoles d'essais en parcelles de taille limitée, ont besoin d'être réexaminées par une observation plus généralisée ; c'est le cas de l'effet de la date de semis, qui prend dès à présent une valeur encore plus affirmée que celle se référant à la seule expérimentation dans les centres d'essais.

Par ailleurs, d'autres facteurs ne peuvent guère entrer dans un protocole expérimental, mais ressortent facilement d'une enquête : par exemple, ici, l'effet de la dimension des champs, objet quasi impossible à expérimenter en station ou en réseau d'essais traditionnels. Ainsi, les domaines d'investigation par expérimentation et par enquête, s'ils se recouvrent, peuvent confirmer leur objet commun, et, s'ils ne se recouvrent pas, peuvent compléter l'information avec profit.

Les défauts apparaissant dans la présente enquête se rattachent d'abord au modèle de la fiche elle-même, telle qu'elle fut progressivement établie par les investigateurs, et qui pourra être améliorée en connaissance de cause. Ils tiennent ensuite à la façon fragmentaire dont les fiches furent remplies par beaucoup d'encadreurs, peu familiarisés avec les exigences de la procédure d'enquête. On a déjà souligné la carence géné-

rale des réponses au poste « rendement » : un effort spécial devrait être demandé aux chefs de secteurs et aux responsables de l'encadrement pour obtenir des informations précises, lesquelles pourraient servir également aux estimations de production effectuées traditionnellement par d'autres méthodes, par les Sociétés de mise en valeur pour chaque zone d'usage. Mais les plus graves défauts sont la disparité et la variation annuelle du volume des contributions à l'enquête par région, causes les plus manifestes des difficultés rencontrées dans l'exploitation statistique des données. Si une enquête du même genre devait se poursuivre, il serait bon d'obtenir, de la part du personnel d'encadrement, l'exécution d'un quota donné de fiches d'observation pour chaque secteur

cotonnier, en conservant un volume de renseignements stable d'année en année, conformément à un plan d'échantillonnage préétabli. De plus, l'élaboration du protocole d'enquête et du questionnaire aurait intérêt à être réalisée en liaison avec les biométriciens chargés de l'exploitation ultérieure des données.

Il demeure que la présente étude, dont la réussite n'est que partielle, permettra d'accorder plus étroitement les objectifs visés par une enquête du type agricole avec les moyens demandés pour sa bonne réalisation, dans le but d'éviter avant tout les questionnaires démesurés, hors de la portée des cultivateurs, et les réponses inconsistantes qui en découlent alors obligatoirement.

### REMERCIEMENTS

Le Ministère du Développement rural a contribué largement à ouvrir les possibilités des enquêtes régionales ; nous lui sommes particulièrement redevables, ainsi qu'au Ministère de la Recherche qui a accordé son soutien au programme concerné par ces investigations.

Les questionnaires furent diffusés, puis collectés grâce à l'attention portée par les Directions des

Offices régionaux de développement à nos requêtes ; nous remercions tous ceux qui, sur le terrain, d'année en année, ont accepté de contribuer à notre enquête, notamment les chefs de secteurs, les moniteurs et les cultivateurs qui ont bien voulu fournir, transmettre et classer des renseignements d'une manière aussi précise que possible.

### BIBLIOGRAPHIE

1. DESMIDTS M., 1975. — Mesures prophylactiques préconisées contre la phylodie du cotonnier. *Rapport présenté au Comité de la Recherche Agronomique de Haute-Volta [Ouagadougou, 1975]*.
2. HUGHES N., 1970. — Méthodes par segmentation et typologie. *Bordas, Paris*.
3. GIANNOTTI J. et R. DELATTRE, 1972. — Une nouvelle approche de l'étude épidémiologique d'une phylodie, la virescence florale du cotonnier : culture sélective de mycoplasmes extraits de quelques plantes et insectes homoptères de la biocénose. *Cot. Fib. trop.*, 27, 4, 371-378.

### SUMMARY

Regional surveys of cotton fields affected by floral virescence (VFC) have been carried out in Upper Volta for 9 years. After critical examination of the data obtained, the card files made were processed by a computer, using two different programs : a multiple regression analysis (MR) program and segmentation analysis (SA) program.

The main components analysed were : sowing or emergence dates, dimensions of fields, number of treatments, fertilisers, previous crops, nature of pesticides, etc.

The results are given in :

- tables of frequencies of various simple or crossed factors ;
- progressively simplified regression formulas ;
- a study of the quality factors ;
- an iterative dichotomy table.

Only the variable resulting from the « percentage of plants per hectare », unlike that resulting from the fields, can be used.

Interesting correlations have been established between the disease level and the first three factors mentioned, in MR analysis ; the relationships revealed

by SA include additional factors, and also provide refinements of the results obtained by MR analysis. Classification in order of importance does not give the same result for both analyses. For exceptional years, SA shows that the « dimensions of the field » action changes when the sowing date is late.

The advantages or disadvantages of the two data processing methods are stated. The possible complementarity of the two investigatory methods, conventional experimentation and collection of data by regional surveys, is emphasized. The latter procedure makes it possible to study certain problems, for example, those relating to the dimensions of fields, which cannot be done by present experimental methods.

As regards the factors studied concurrently by the two methods in Upper Volta, there is good agreement between the conclusions reached, in particular in regard to the incidence of an early sowing date on the disease level. Nevertheless, the practical recommendations to sow early is maintained, as this is the basis of a good productivity potential, so long as the virescence level is not exceptionally high.

## RESUMEN

Durante nueve años, se llevaron a cabo en Alto Volta encuestas regionales en los campos de algodones atacados por la virescencia floral (VFC). Después de examen crítico de los datos recogidos, los ficheros constituidos de esta manera fueron explotados por dos programas informáticos diferentes: regresión múltiple (R.M.) y análisis por segmentación (A.S.).

Los principales elementos analizados son: fecha de siembra o de brote, dimensión de los campos, número de tratamientos, abónos químicos, cultivo precedente, naturaleza de los pesticidas, etc.

Los resultados son enunciados mediante:

- cuadros de frecuencia de diversos factores, simples o cruzados;
- fórmulas de regresión progresivamente simplificadas;
- un estudio de los factores cualitativos;
- una tabla dicotómica iterativa.

Sólo la variable resultante de los « porcentajes de plantas por hectarea » es explotable, pero no fue posible estudiar la del rendimiento.

Dos correlaciones interesantes se establecen entre el porcentaje de enfermedad y los tres primeros factores citados en R.M.; las relaciones evidenciadas

por A.S. comprenden factores suplementarios y aportan también matices complementarios a los resultados clarificados en R.M. La clasificación por orden de importancia no es la misma en un análisis que en el otro; para años excepcionales, el A.S. muestra que la acción « dimensión del campo » está modulada diferentemente cuando la fecha de siembra es tardía.

El interés o los inconvenientes respectivos de ambos enfoques informáticos son evocados; se apunta la eventual complementariedad de ambas técnicas: experimentación clásica y recogida de datos mediante encuestas regionales; este último proceso permite abordar ciertos problemas (relación con la dimensión de los campos, por ejemplo), que no pueden ser tratados en experimentación corriente.

En cuanto a los factores estudiados en el Alto Volta concurrentemente por las dos técnicas, acusan una concordancia correcta en las conclusiones, en particular a propósito de la incidencia de la fecha de siembra precoz sobre el porcentaje de enfermedad. Sin embargo, la recomendación práctica de efectuar siembras precoces se mantiene, puesto que es el fundamento de un buen potencial de productividad, por lo tanto que el porcentaje de virescencia no sea particularmente elevado.